Human Activity Recognition with Smartphone Sensor Data

1. Introduction

Personalized health care has been a hot topic in the field since the evolution of IoT and Big Data. Real-time sensor data and analysis enable us to provide personalized and preventive medicine, as well as monitoring, in the framework of Internet-of-Humans [1]

In this study, we try to identify 6 different activities based on data from sensors that are embedded in waist-mounted smartphones. We used the Human Activity Recognition (HAR) data which contains recordings of 30 volunteers, within age range of 19-48 years, performing activities of daily living (ADL): Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing and Laying [2].

Exploratory analysis and predictive modeling are conducted to recognize distinctive activities. Two classification models are compared in Section 4, multinomial logistic regression and XGBoost. Upon initial experiment, multinomial logistic regression and XGBoost can achieve total accuracy of 93.64% and 91.34% separately.

2. Data Collection and Preparation

2.1 Data Source

The data was collected from experiments on a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (walking, walking upstairs, walking downstairs, sitting, standing, and laying) wearing a smartphone (Samsung Galaxy S II) on the waist [2].

Using the accelerometer and gyroscope embedded in the sensor, 3-axial linear acceleration and 3-axial angular velocity were captured at a constant rate of 50Hz. The data was manually labeled based on recorded experiment videos. Time domain signals (prefix 't') were pre-processed by applying a median and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (*tBodyAcc-XYZ* and *tGravityAcc-XYZ*) using another low pass Butterworth filter with a corner frequency of 0.3 Hz [2].

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (*tBodyAccJerk-XYZ* and *tBodyGyroJerk-XYZ*). And the magnitude of these three-dimensional signals were calculated using the Euclidean norm (*tBodyAccMag*, *tGravityAccMag*, *tBodyAccJerkMag*, *tBodyGyroMag*, *tBodyGyroJerkMag*) [2].

Finally, a Fast Fourier Transform (FFT) was applied to some of these signals producing *fBodyAcc-XYZ*, *fBodyAccJerk-XYZ*, *fBodyGyro-XYZ*, *fBodyAccJerkMag*, *fBodyGyroMag*, *fBodyGyroJerkMag* (prefix 'f' indicates frequency domain signals) [2].

These signals were used to estimate variables of the feature vector for each pattern:
'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions. Furthermore, summary statistics of each feature were calculated as additional features aiming at providing a more detailed and complete description of the actual behavior [2].

2.2 Data Overview and Preparation

The final dataset contains 561 features for 10,299 records from 30 different experiment subjects. To build a predictive model for identifying different activities later on, we first split the data into training and testing set. About 80% of data were used as the training data and the rest 20% as testing set. In order to approximate 80% of the total records as close as possible while do not have the same subject appearing in both training and testing set, we determined which subjects can be grouped together by calculating the cumulative sum of the records for them. As a result, by leaving subjects 2, 5, 7, 8, 9, 10, 11 in the testing set and put other subjects in the training set, we can result in an 80-20 split.

3. Exploratory Analysis

Preliminary analysis was done at the beginning in order to have a better understanding of the general data distribution. For simplicity, only some core features were analyzed in this stage to explore potential relationship with different activities. Core features include *tBodyAcc-XYZ, tGravity-XYZ, tBodyAccJerk-XYZ, tBodyGyro-XYZ, tBodyGyroJerk-XYZ, tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag,* and *tBodyGyroJerkMag*. Among all kinds of summary statistics, we picked the mean value to further explore.

Figure 1 shows the correlation heatmap of selected core features. The color indicates the correlation between every pair of two features, ranging from -1 to +1, where dark blue means a perfectly negatively linear relationship while dark red refers to a perfectly positively linear relationship.

From Figure 1, we can see high correlations among the last few features (*tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag*) and also some relationships between those features and *tBodyAcc-XYZ*. It is also obvious to see some relationships within the same sensor data (e.g. *tBodyAcc*) in XYZ coordinates.

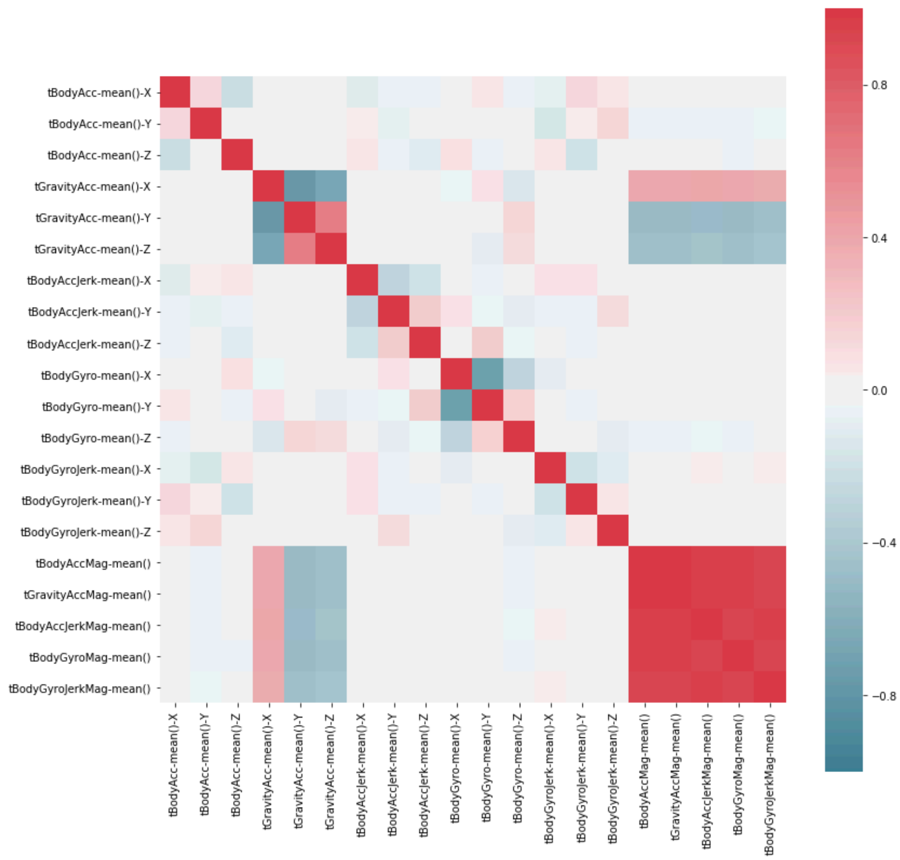

Figure 1. Correlation Heatmap of Core Features

Furthermore, we look at how different activities vary in some sensor data in 3-dimensional space (X-Y-Z directions).

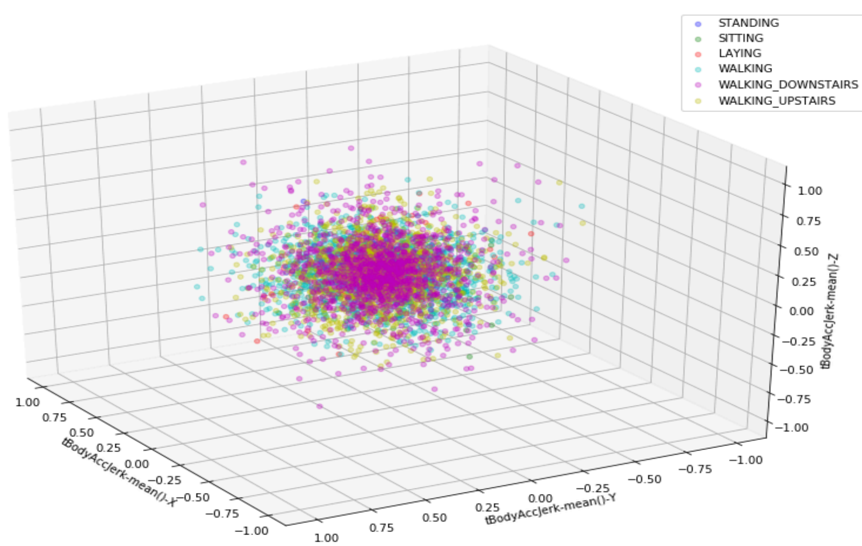

Figure 2. Records of Different Activities in 3D Space (tBodyAccJerk)

It is hard to separate different activities based on some features (Figure 2). But some other sensor data clearly show the distinctions among different behaviors (Figure 3). Hence, it is possible to identify different activities based on the HAR data.

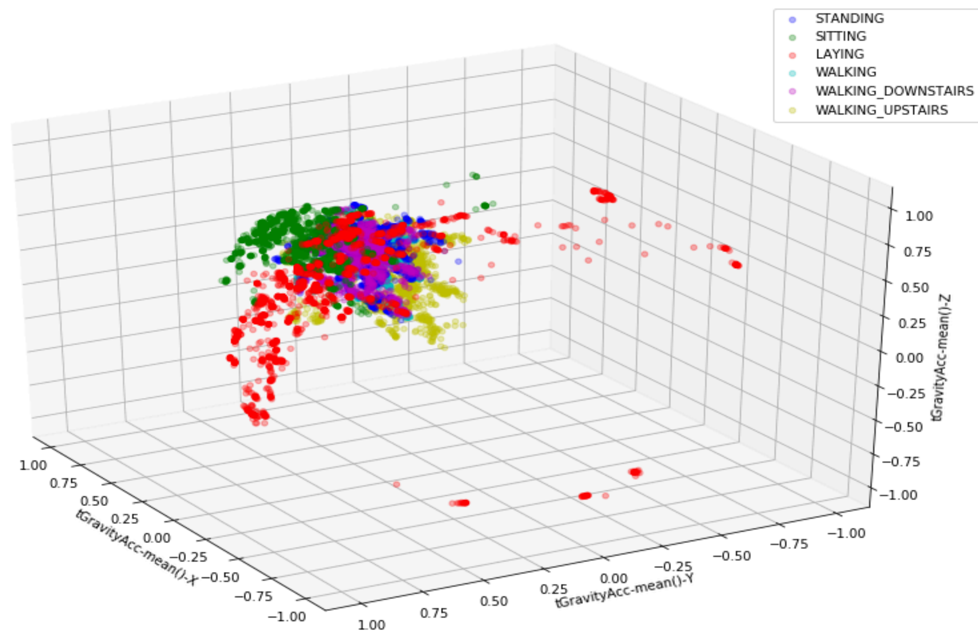

Figure 3. Records of Different Activities in 3D Space (tGravityAcc)

4. Methods

In this study, we compared two different models, multinomial logistic regression and XGBoost.

4.1 Models

4.1.1 Multinomial Logistic Regression

Multinomial logistic regression is a statistical model that uses logistic function to model the probability of several classes or events existing or happening [3]. Newton's method was used to estimate the parameters of the model subject to minimizing the multinomial loss (cross-entropy loss) across the entire probability distribution.

$$\text{Cross Entropy Loss} = J(\theta) = \sum_{i=1}^{m} -y^{(i)} \log h_\theta(x^{(i)}) - (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))$$

4.1.2 XGBoost

XGBoost is a tree-based model that is an ensemble of decision trees using gradient boosting and it is computed in parallel which yields results fast and accurate [4]. The model tries to minimize the softmax objective function.

4.2 Feature Selection

To reduce the complexity of the model, we selected a subset of features based on the feature importance from XGBoost model. The importance of each feature was evaluated based on the average entropy gain across all splits in the tree model when the feature is used.

The purpose is to select the least number of features that can maintain a descent model accuracy (80% or 90%). Features were selected from the most important one to the least important.

4.2.1. Dimensionality Reduction

In addition to the experiment of simple feature selection, we also applied principal component analysis [5] on the original features and aimed to achieve the same model performance (90% accuracy on testing set) with even fewer features (principal components).

5. Results

5.1 Multinomial Logistic Regression

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	371
SITTING	0.88	0.89	0.88	341
STANDING	0.90	0.88	0.89	353
WALKING	0.90	0.99	0.94	384
WALKING_DOWNSTAIRS	0.98	0.98	0.98	305
WALKING_UPSTAIRS	0.97	0.88	0.93	337
avg / total	0.94	0.94	0.94	2091

Figure 4. Classification Result of Multinomial Logistic Regression

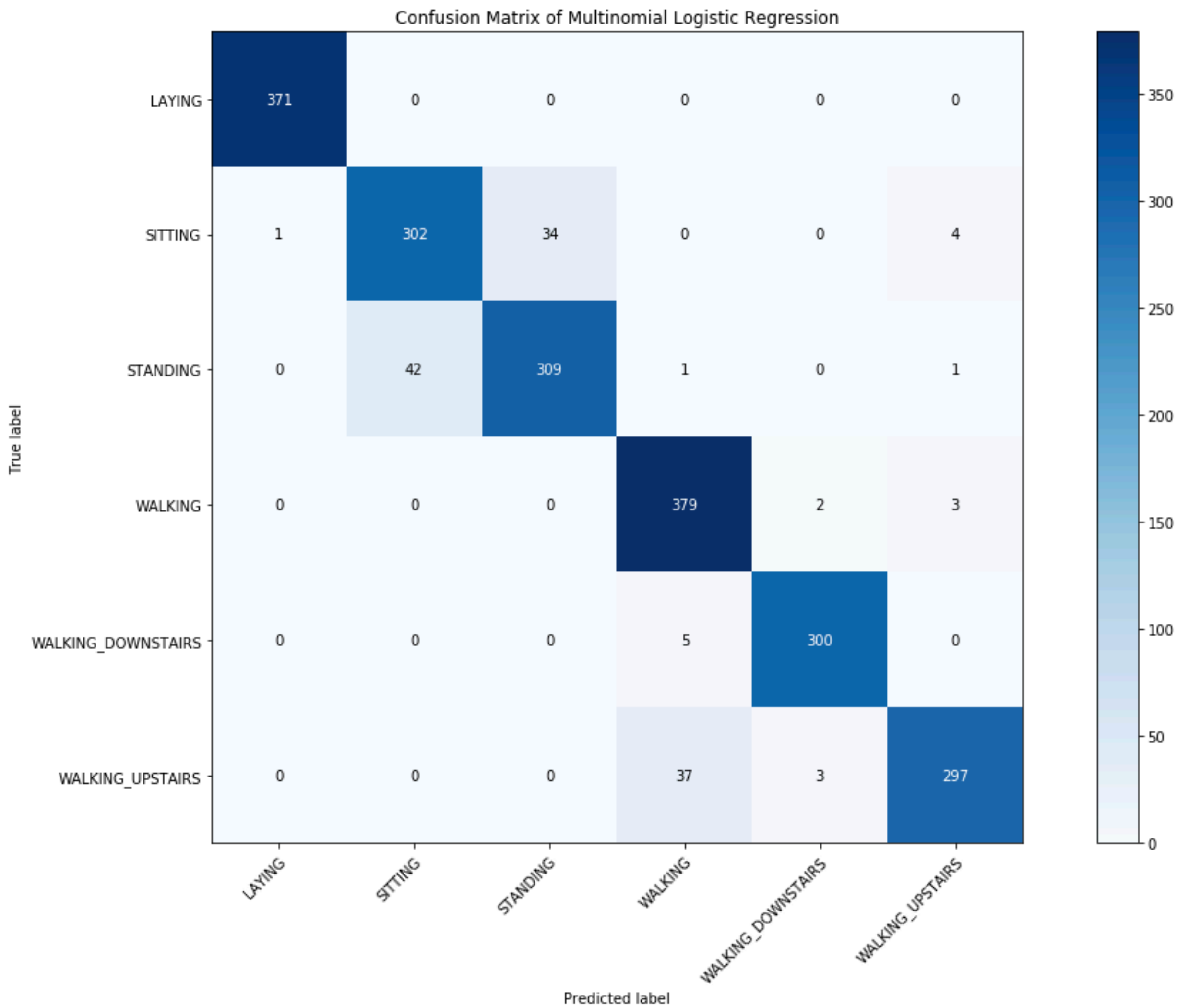

Figure 5. Confusion Matrix of Multinomial Logistic Regression

Based on Figure 4 and Figure 5 above, multinomial logistic regression has descent performance on the evaluation of testing set. It has overall accuracy of 93.64% (approximated to 94% on the figure), while it correctly predicted all the records of LAYING activity. Having such performance indicates that the problem is linearly separable. The multinomial logistic regression has the least performance on predicting the activity of SITTING, where it only correctly predicts 88% of the records from the SITTING activity.

5.2 XGBoost

Both Figure 6 and Figure 7 down below show a good model performance of XGBoost on the testing set. The overall accuracy of XGBoost model is 91.34% (approximated to 92% on the figure), and it also correctly predicted all the records of LAYING activity.

Similar to the multinomial logistic regression, XGBoost also has the least performance on predicting the activity of SITTING, where it only correctly predicts 78% of the SITTING activities in the test set.

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	371
SITTING	0.78	0.86	0.82	341
STANDING	0.86	0.77	0.81	353
WALKING	0.92	0.97	0.95	384
WALKING_DOWNSTAIRS	0.94	0.99	0.97	305
WALKING_UPSTAIRS	0.98	0.89	0.93	337
avg / total	0.92	0.91	0.91	2091

Figure 6. Classification Result of XGBoost

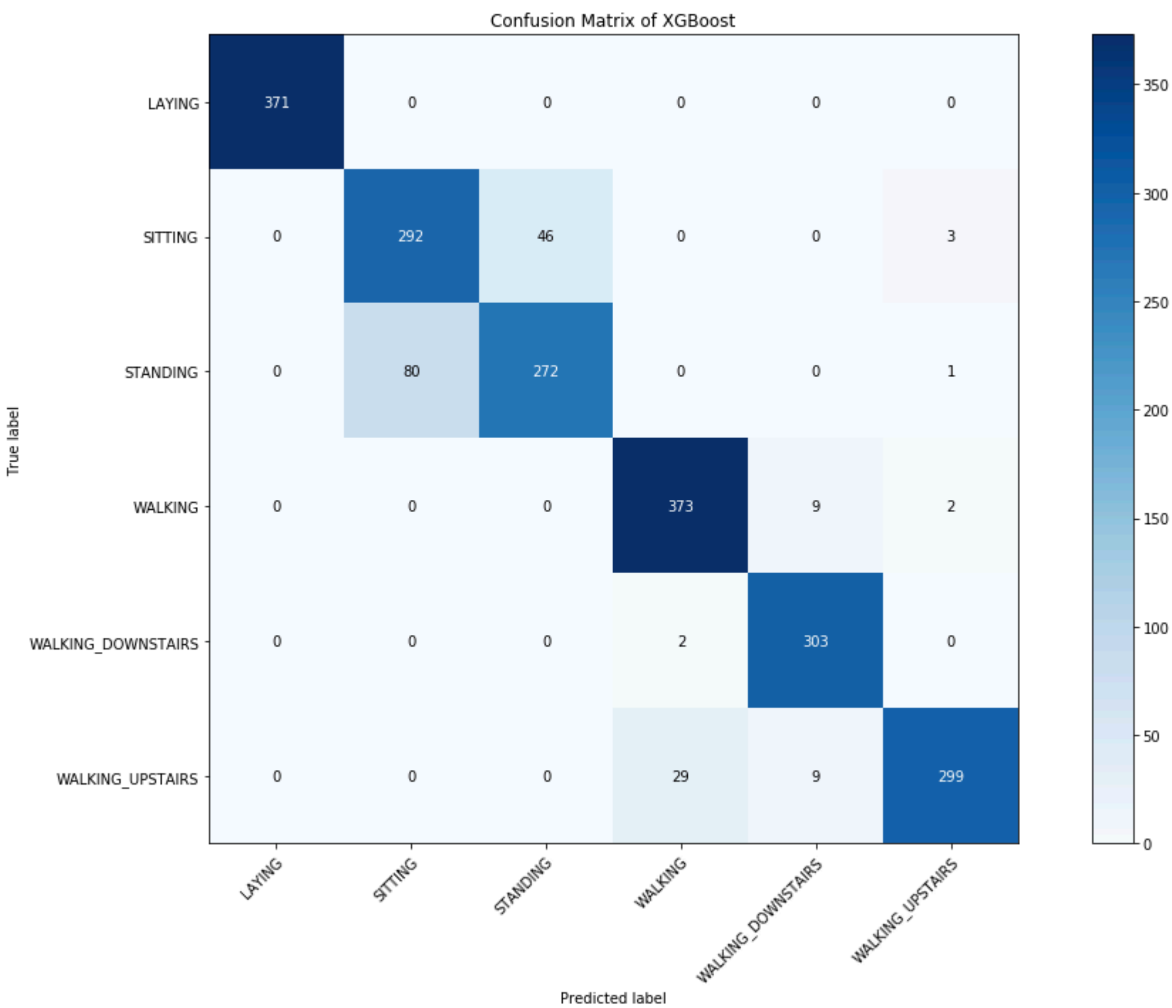

Figure 7. Confusion Matrix of XGBoost

5.2.1 XGBoost with Reduced Number of Features

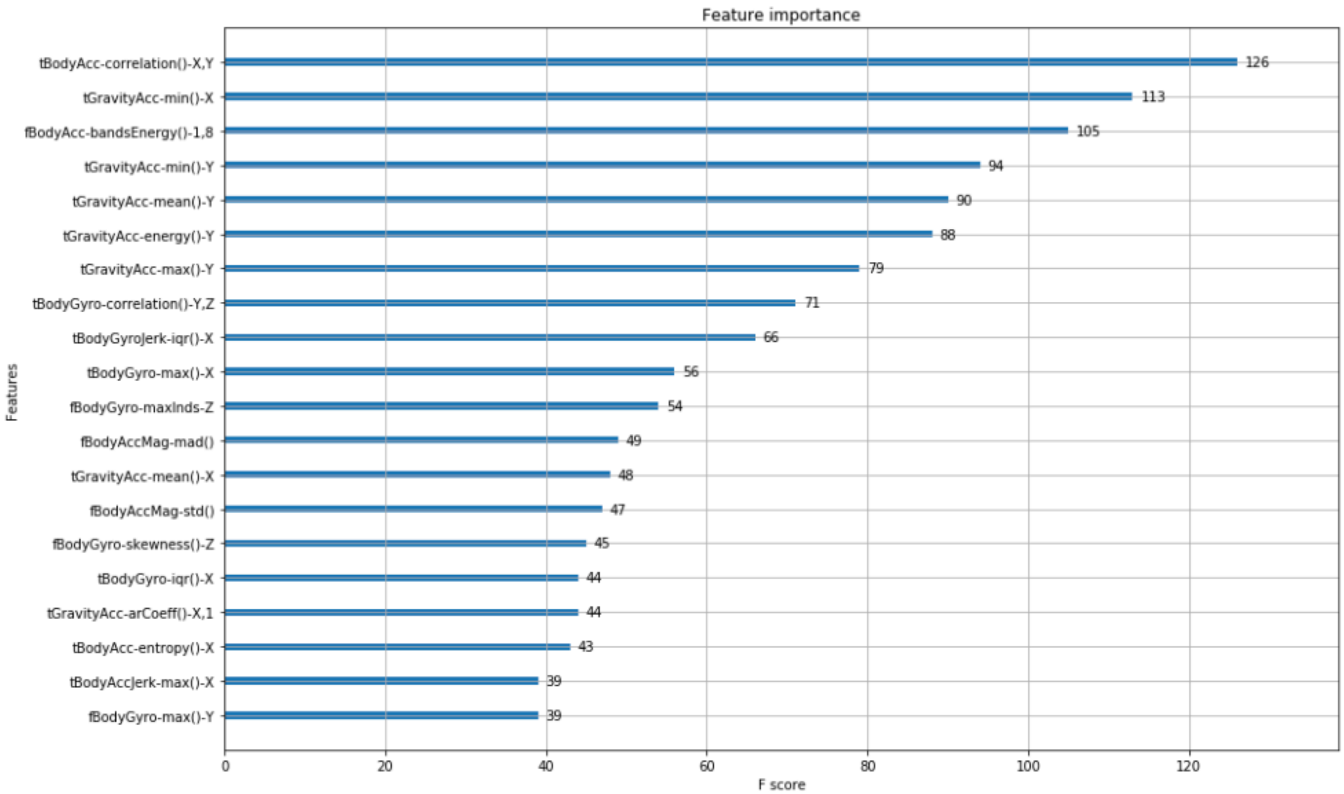

Figure 8. Feature Importance of XGBoost

As mentioned in Section 4.2, a subset of features was selected based on the feature importance (Figure 8). As a result, we can maintain 80% model accuracy with only 10 original features and achieve 90% model accuracy with 53 original features (Figure 9).

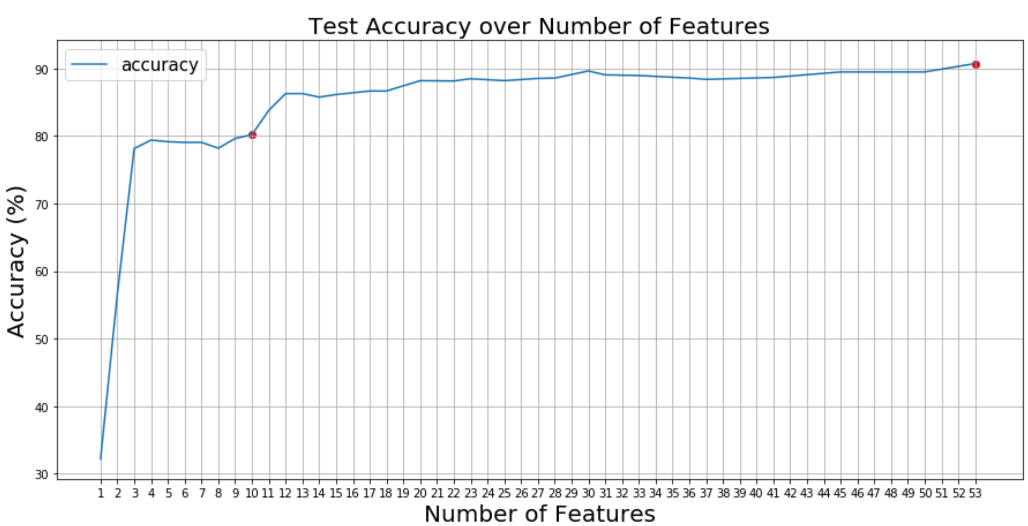

Figure 9. Change of Test Accuracy on XGBoost with Original Features

5.2.2 XGBoost with Principal Component Analysis (PCA) Components

Based on the Figure 10 below, we do achieve the same model performance, test accuracy of 90%, with 20 less features. Number of features in the model dropped from 53 original features to 33 PCA components.

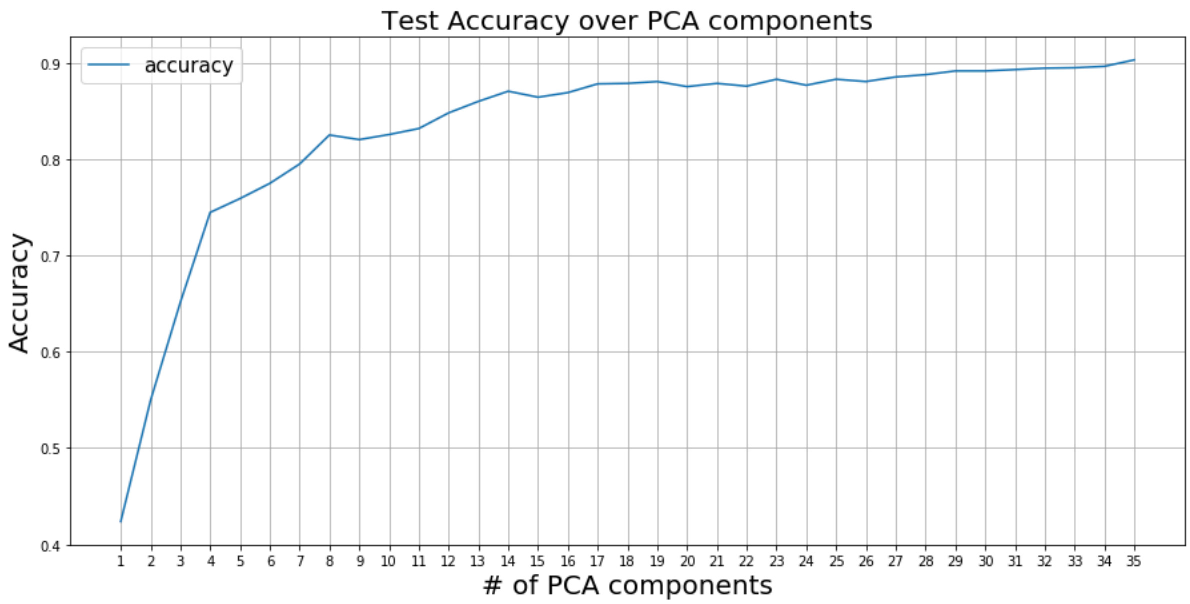

Figure 10. Change of Test Accuracy on XGBoost with Principal Component Analysis

6. Conclusion

Based on the Human Activity Recognition (HAR) data [2], we can accurately predict (About 90% total accuracy) 6 different daily activities using machine learning methods. From this study, we observe that sensor data on horizontal body acceleration (*tBodyAcc-correlation()-X,Y*) and horizontal gravity acceleration (*tGravityAcc-min()-X*) are very important at separating different activities.

In addition, among all ADLs (Activity of Daily Living), laying is the one with the highest prediction accuracy, which also aligns with our intuition since laying is the activity that has the least movement compared to others.

References

1. Hans Lehrach, Adrian M. Ionescu. “The Future of Health Care: deep data, smart sensors, virtual patients and the Internet-of-Humans”. 2016
2. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013. https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones#
3. "Logistic regression." Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc. 22 July 2004. Web. 10 Aug. 2004. https://en.wikipedia.org/wiki/Logistic_regression
4. XGBoost Documentation. https://xgboost.readthedocs.io/en/latest/
5. "Principal component analysis." Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc. 22 July 2004. Web. 10 Aug. 2004. https://en.wikipedia.org/wiki/Principal_component_analysis